Online Shop Customer Sales

MOHAMMED BAJODAH

Analysis Framework

- [What] I obtained sales data from an online shop with 60 thousands customer data.
  - The issue is lack of understanding in consumer behavior engaging with E-commerce
- [How] I will use regression analysis and cluster analysis to understand the optimal approach to online shopping. This provides value to stakeholder to increase revenues.
- [Why] this an important area to advance for retailers and e-commerce as using data to provide insights can improve business revenue cycles and deliver a larger value to stakeholder.

Data Description

Here is a brief description of each variable:

I. **Customer_id:** A unique identifier for each customer.

II. **Age:** The age of the customer.

III. **Gender:** A binary variable where 0 represents male and 1 represents female.

IV. **Revenue_Total:** Total sales revenue by the customer.

V. **N_Purchases:** The number of purchases made by the customer to date.

VI. **Purchase_DATE:** The date of the latest purchase made by the customer.

VII. **Purchase_VALUE:** The value of the latest purchase made by the customer in euros.

VIII. **Pay_Method:** A categorical variable indicating the payment method used by the customer. The - - categories are digital wallets, card, PayPal, and other.

IX. **Time_Spent:** The time spent by the customer on the website in seconds.

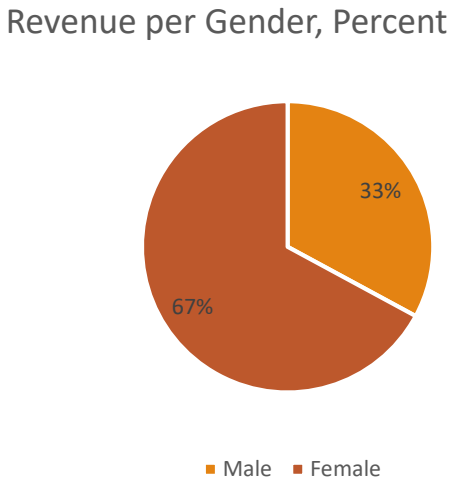
X. **Browser:** A categorical variable indicating the browser used by the customer. The categories are Chrome, Safari, Edge, and other.

XI. **Newsletter:** A binary variable indicating whether the customer is subscribed to the newsletter or not.

XII. **Voucher:** A binary variable indicating whether the customer has used a voucher or not.

Various analysis can be extracted from the description on the left using *revenue_total* variable as a dependent variable where correlation with other variables (RHS)

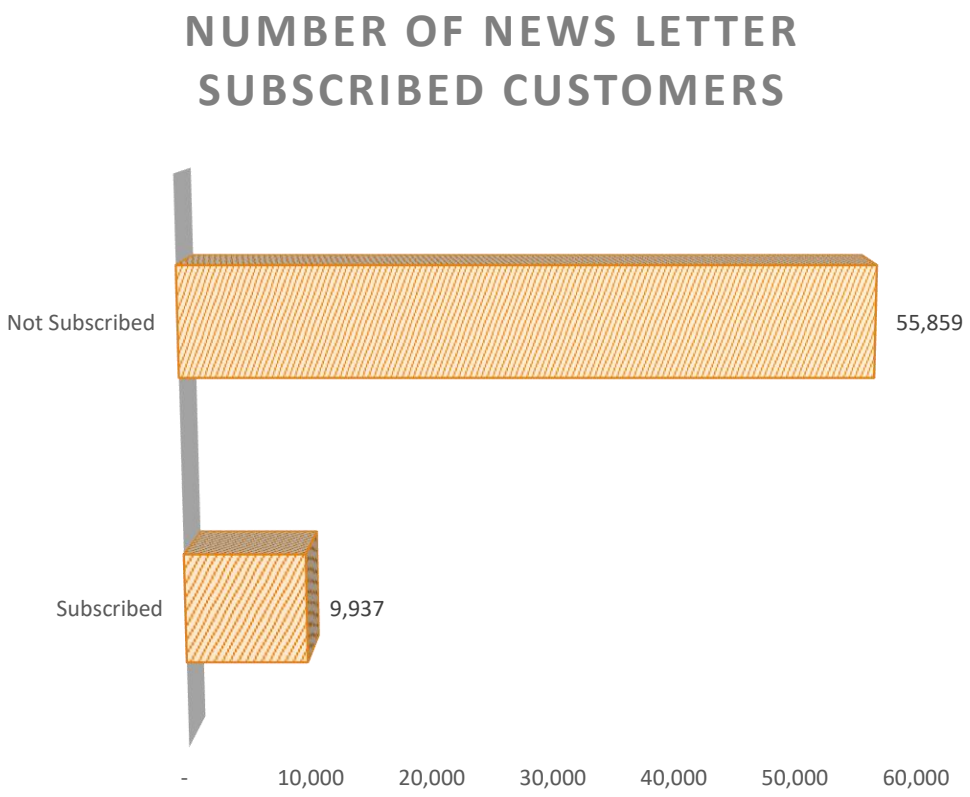
Diagnostics and Visualization (1/4)



		Male		Female
Number of Purchase		86,242		176,447
Total Revenue	$	600,161	$	1,224,555
Rev. per Purchase		7		7

The data suggest that two thirds of purchases are made by females with a total revenue of 1.2 million generated. There is no evidence to believe that females spending per purchase is higher than males.

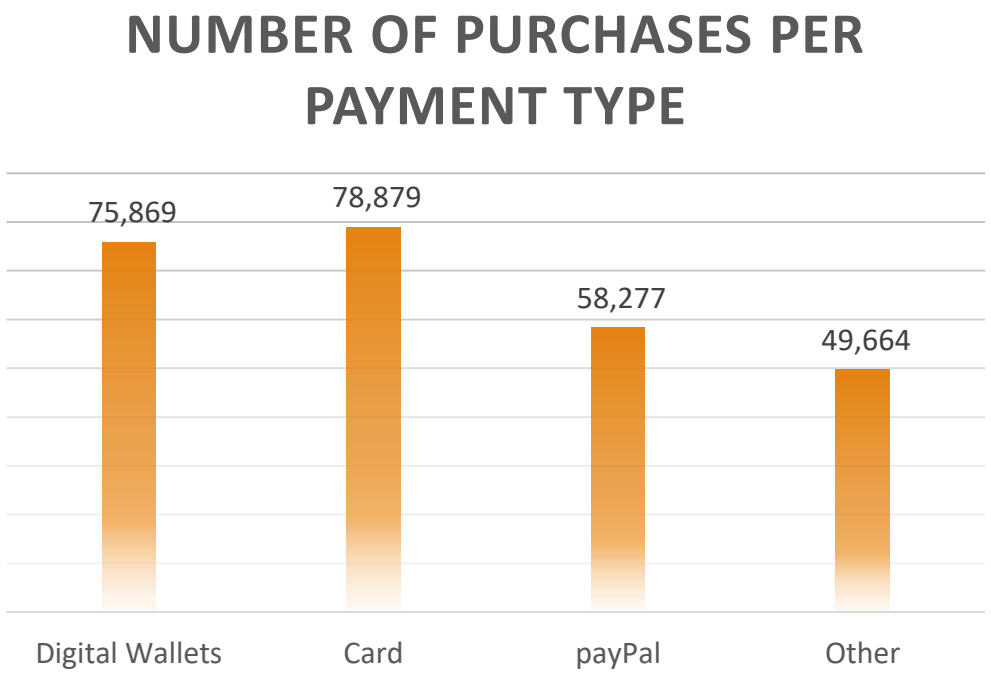
Diagnostics and Visualization (2/4)

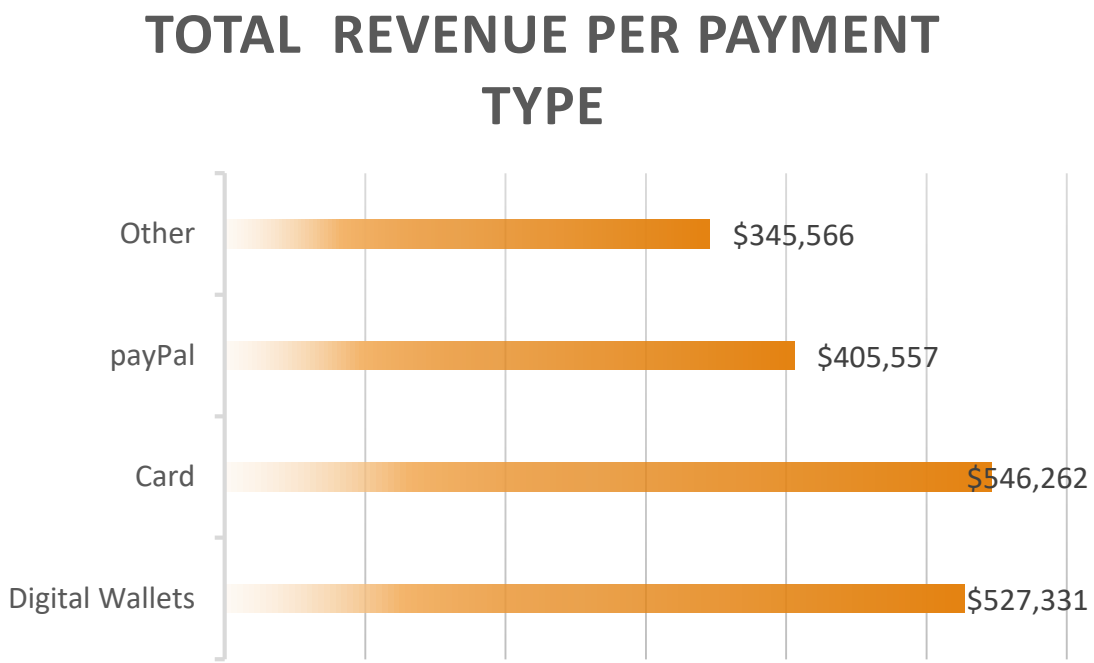


Age Group	Number of customers	Number of purchase	Total Revenue
16 - 25	13,539	54,036	$ 375,562.3
26 - 35	13,632	54,989	$ 376,502.2
36 - 45	13,738	54,739	$ 383,389.7
46-55	13,922	55,228	$ 384,786.3
56 and above	10,965	43,697	$ 304,475.7

The dataset suggests that customer purchases measured in both revenue or frequency of purchases are similar in age groups (except the older population). In addition, the newsletter subscription are not leading to more sales..

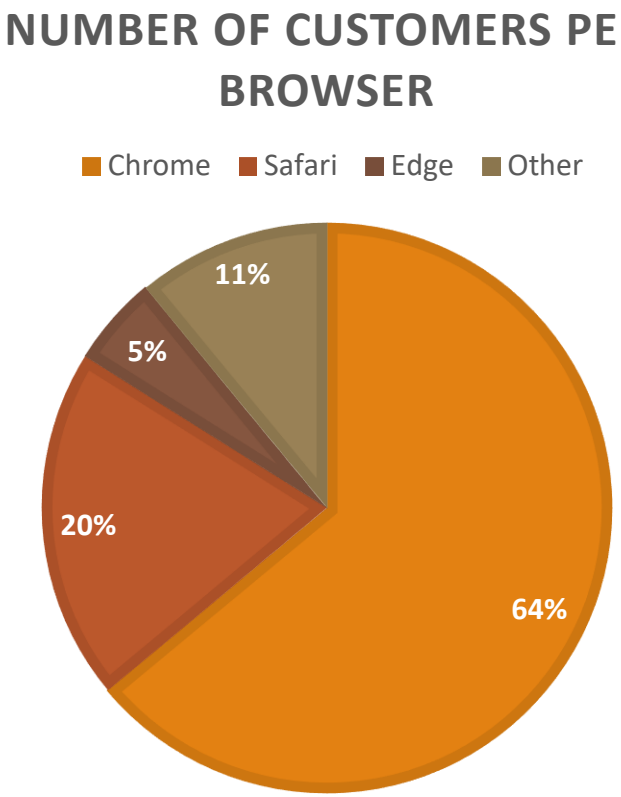
Diagnostics and Visualization (3/4)



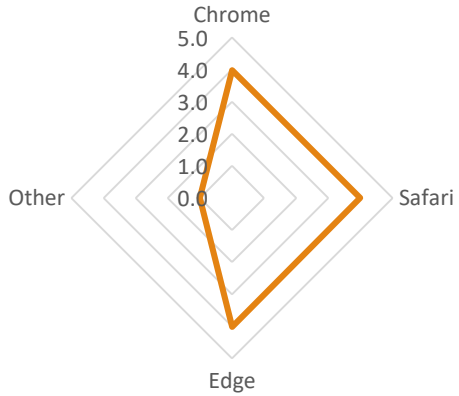


The most common payment is by Card and followed by Digital Wallets. The revenues generated from cstomers using Card and Digital Wallets are much higher than other payment platforms

Diagnostics and Visualization (4/4)



frequancy of purchases (i.e. returning customers)



Most customers use Google Chrome to shop online and buy items. However, on average, all browser face the same frequency of 4 purchases per user except for those who use “other” only purchase it at one time..

Regression Analysis 1/3

Having diagnosed the data in the previous work, now we run regression analysis to predict the revenue from each customer using the following dependent variables:

1. Age
2. Gender
3. Payment Type
4. Time spent on the website

To give top management a view on how to structure their website to optimize revenues from their customers.

Customer_id	Age	Gender	Pay_Method	Time_Spent	Revenue_Total
504308	53	0	1	885	$ 45
504309	18	1	2	656	$ 36
504310	52	1	0	761	$ 11
504311	29	0	1	906	$ 54
504312	21	1	1	605	$ 57
504313	55	0	1	364	$ 14
504314	17	1	0	654	$ 31
504315	30	1	3	1011	$ 8
504316	51	0	0	312	$ 18
504317	63	1	3	828	$ 19

Top 10 rows of the data base

Regression Analysis 2/3

Regression Statistics	
Multiple R	0.006859552
R Square	4.70535E-05
Adjusted R Square	-1.37423E-05
Standard Error	14.94198053
Observations	65796

ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	4	691.1857823	172.7964456	0.773959922	0.541921168			
Residual	65791	14688681.7	223.2627822					
Total	65795	14689372.89						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	27.75937905	0.242552895	114.4467028	0	27.28397537	28.23478273	27.28397537	28.23478273
Age	9.01775E-05	0.004213406	0.021402509	0.982924637	-0.008168099	0.008348454	-0.008168099	0.008348454
Gender	-0.030631677	0.124015726	-0.246998327	0.80491029	-0.273702506	0.212439152	-0.273702506	0.212439152
Pay_Method	0.080114583	0.053841335	1.487975421	0.136762168	-0.025414436	0.185643601	-0.025414436	0.185643601
Time_Spent	-0.000190828	0.000209669	-0.910140874	0.362751555	-0.000601779	0.000220123	-0.000601779	0.000220123

Regression Analysis 3/3

The regression analysis used in this study yields the following:

1. Evidence suggests that male and female shoppers have an equal probability of spending an average of 27 dollars for their purchases.

2. The payment methods available on the website does not optimize purchase value

3. Time spent on the website does not correlate positively with the frequency, or ticket size of the purchase.

4. Older customers tend to have a larger ticket size then those who are younger. However, the results are not statistically significant

Opportunities for Management to take evidence-based approach in optimizing revenue

Based on the results above, I recommend the following:

1. Focus marketing campaigns on older people who know what they want already (results indicate they don't want to spend more time on the website)

2. No further investment in payment integrations is needed as indicated by the analysis

3. Further work in data collection (i.e. what do they buy) and statistical modeling is needed to understand consumer behavior (limitation on this dataset)